

Indicateurs SIDEP à partir des tables agrégées

12/05/2022

Modification de la mise à disposition des indicateurs issus de SIDEP

Afin de permettre la mise à disposition des indicateurs épidémiologiques issus de SIDEP, de nombreux scripts vont lire la base de données « brute ». Cette méthode nécessite de recalculer tous les jours des indicateurs complexes (définition patient incident), et de découper les données sous formes de nombreuses agrégations géographiques (nationale, régionale, départementale, EPCI, IRIS) et temporelles (quotidien, hebdomadaire et 7 jours glissants).

La production de ces indicateurs calculés sur l'ensemble de l'historique chaque jour est une forte contrainte sur les infrastructures de SpF, d'où la nécessité de construire un modèle de données plus efficient.

Dans ce but, des bases de données pré-agrégées selon la « définition patient » et selon différentes échelles géographiques ont été développées afin d'alléger les temps de traitement pour la mise à disposition des indicateurs épidémiologiques. Ces bases de données permettent de mettre à disposition des données sur l'ensemble de la période, et sont actualisées quotidiennement pour les 15 derniers jours (sachant que 95% des données arrivent dans un délai de 3 jours entre la date de prélèvement et l'intégration dans SIDEP).

Ce mode de calcul permettra de sécuriser le processus de production des indicateurs SIDEP et de minimiser les temps de traitements. Pour cela, le format des fichiers SIDEP ainsi que leur contenu est modifié afin d'optimiser le nombre de fichiers mis à disposition.

Ces modifications ne changent pas le mode de calcul des différents indicateurs mis à disposition (TI, TP et TD).

Le nom des fichiers :

La nomenclature des fichiers est modifiée :
Tous les fichiers issus de SIDEP commencent par l'abréviation « sp » suivis par l'échelon géographique, le pas de temps et les variables d'agrégations (age, sexe...).

Par exemple :

- sp_dep_7j_cage10 → contient les données au niveau départemental pour les semaines glissantes et les classes d'âge de 10 ans.
- sp_fra_heb_cage_scol → contient les données au niveau national pour les semaines calendaires et les classes d'âge scolaires

Abréviation	Signification
Sp	SIDEP
Fra	France
Reg	Région
Dep	Département
Iris	Numéro iris
Epci	Code de l'epci
Com	Code de la commune (code INSEE)
7j	Semaines glissantes

Heb	Semaines calendaires
Jour	Quotidien
Age	Age
Cage10	Classe d'âge de 10 ans
Cage15	Classe d'âge de 15 ans
Cage65	Classe d'âge de 65 ans
Cage_scol	Classe d'âge scolaire
Cl_age90	Borne supérieure de la classe d'âge de 10 ans
Cl_age18	Borne supérieure de la classe d'âge scolaire
Sexe	Sexe (F, M, NA)

Le contenu des fichiers :

La quantité de fichiers produits diminue mais le contenu de chaque fichier augmente.

Les fichiers contiennent dans l'ordre suivant :

- Une colonne pour l'échelon géographique :
 - o Fra = France
 - o Reg= code de la région
 - o Dep= code du département
 - o IRIS= code IRIS
 - o Com = Code de la commune (code INSEE)
 - o Epci = Code de l'EPCI
- Une colonne pour le pas de temps :
 - o Sg = semaine glissante **ou**
 - o Semaine = semaine calendaire **ou**
 - o Date = date du jour
- Une colonne pour les variables d'agrégations :
 - o Cage_scol= [00-03), [03-06), [06-11), [11-15), [15-18), [18-Inf) **ou**
 - o Cage10= [00-10), [10-20), [20-30), [30-40), [40-50), [50-60), [60-70), [70-80), [80-90), [90-Inf) **ou**
 - o Cage15= [00-15), [15-45), [45-65), [65-75), [75-Inf) **ou**
 - o Cage65= [00-65), [65-Inf)
 - o Sexe= M, F
- Une colonne pour chaque effectif :
 - o Pop= population
 - o P = patients testés positifs
 - o T= nombre de patients testés
- Une colonne pour chaque indicateur :
 - o Ti = taux d'incidence
 - o Tp= taux de positivité
 - o Td= taux de dépistage
 - o Ti_classe (uniquement dans les fichiers iris,epci,commune) = La classe du taux d'incidence
 - o Td_classe (uniquement dans les fichiers iris,epci,commune) = La classe du taux de dépistage

- Tp_classe (uniquement dans les fichiers iris,epci,commune) = La classe du taux de positivité

Si le fichier contient une colonne avec la population, il contiendra les trois indicateurs Ti, Tp, Td. Sinon, il contiendra uniquement le taux de positivité.

Les valeurs manquantes pour les variables d'agrégation (y compris l'âge) sont renseignées comme <NA> dans tous les fichiers. Dans le cas particulier de l'âge, il sera sûrement nécessaire de retransformer le type de l'âge (de « string » vers « integer »).

A noter :

- Les fichiers sur le pas de temps hebdomadaires ne seront mis à jour qu'une fois par semaine, le mercredi afin d'avoir des données consolidées sur la semaine précédente.
- La colonne cl_age90 ayant une valeur « 0 » correspond à « tous ages » dans les fichiers qui ne contiennent pas de classe d'âge.

Correspondance avec les fichiers précédents

Nom_ancien format	Nom_nouveau_format	Commentaire
sg_com_opendata	sp_com_7j_cage65_opendata	Le séparateur est un point-virgule et le format des classes des taux d'incidence, dépistage et positivité est modifié*
sg_epci_opendata	sp_epci_7j_cage65_opendata	Le séparateur est un point-virgule et le format des classes des taux d'incidence, dépistage et positivité est modifié*
sg_iris_opendata	sp_iris_7j_cage65_opendata	Le séparateur est un point-virgule et le format des classes des taux d'incidence, dépistage et positivité est modifié*
sg_metro_opendata	sp_metro_7j_cage65	Le séparateur est un point-virgule.
sp_18ans_dep	sp_dep_heb_cage_scol	
sp_18ans_fra	sp_fra_heb_cage_scol	
sp_18ans_reg	sp_reg_heb_cage_scol	
sp_capa_heb_fra		A retirer (uniquement des indicateurs "patients" en open-data)
sp_capa_heb_reg		A retirer (uniquement des indicateurs "patients" en open-data)
sp_capa_heb_dep		A retirer (uniquement des indicateurs "patients" en open-data)
sp_cap_quot_fra		A retirer (uniquement des indicateurs "patients" en open-data)
sp_cap_quot_reg		A retirer (uniquement des indicateurs "patients" en open-data)
sp_cap_quot_dep		A retirer (uniquement des indicateurs "patients" en open-data)
sp_mut_7j_fra	sp_mut_7j_fra	
sp_mut_7j_reg	sp_mut_7j_reg	
sp_mut_7j_dep	sp_mut_7j_dep	
sp_pe_std_heb_dep		A retirer (uniquement des données brutes non redressées en open-data)

sp_pe_std_heb_fra		A retirer (uniquement des données brutes non redressées en open-data)
sp_pe_std_heb_reg		A retirer (uniquement des données brutes non redressées en open-data)
sp_pe_std_quot_fra		A retirer (uniquement des données brutes non redressées en open-data)
sp_pe_std_quot_reg		A retirer (uniquement des données brutes non redressées en open-data)
sp_pe_std_quot_dep		A retirer (uniquement des données brutes non redressées en open-data)
sp_pe_tb_heb_fra	sp_fra_heb_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pe_tb_heb_reg	sp_reg_heb_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pe_tb_heb_dep	sp_dep_heb_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pe_tb_quot_fra	sp_fra_jour_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pe_tb_quot_reg	sp_reg_jour_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pe_tb_quot_dep	sp_dep_jour_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pos_heb_fra	sp_fra_heb_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pos_heb_reg	sp_reg_heb_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pos_heb_dep	sp_dep_heb_cage10	ajout T; TI;TP;TD pré-calculé
sp_pos_quot_fra	sp_fra_jour_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pos_quot_reg	sp_reg_jour_cage10	Suppression de la variable "sexe", non utilisée ; ajout T; TI;TP;TD pré-calculé
sp_pos_quot_dep	sp_dep_jour_cage10	ajout T; TI;TP;TD pré-calculé
sp_ti_tp_7j_18ans_fra	sp_fra_7j_cage_scol	
sp_ti_tp_7j_18ans_reg	sp_reg_7j_cage_scol	
sp_ti_tp_7j_18ans_dep	sp_dep_7j_cage_scol	
sp_ti_tp_7j_fra	sp_fra_7j	ajout T; TI;TP;TD pré-calculé
sp_ti_tp_7j_reg	sp_reg_7j	ajout T; TI;TP;TD pré-calculé
sp_ti_tp_7j_dep	sp_dep_7j	ajout T; TI;TP;TD pré-calculé

***Les classes des indicateurs inclus dans les fichiers infra-départementaux :**

Classes du Taux d'incidence (Ti_classe)	Classes du Taux de dépistage (Td_classe)	Classe du Taux de positivité (Tp_classe)
[000-010), [010-020), [020-050), [050-150), [150-250), [250-500), [500-1000), [1000-Inf)	[000-300), [300-600), [600-800), [800-1000), [1000-1500), [1500-2000), [2000-2500), [2500-Inf)	[00-01), [01-05), [05-10), [10-15), [15-20), [20-25), [25-30), [30-Inf)